

Chapter 9

The Power of Combination Models in Energy Demand Forecasting



Erkan Isikli and Seyda Serdar Asan

Abstract Accurate and reliable energy demand forecasting is essential for sustainable planning, operation, and control of energy systems in our ever-changing world, and it can assist planners and policymakers in making informed decisions to design more efficient power grids, identify cost-effective energy production methods, and optimize energy resource utilization. The related literature emphasizes the need for models that are accurate, fast, and not overly complex to make robust decisions based on forecasts. Time series forecasting offers potential benefits by reducing risks and costly infrastructure upgrades. Particularly, combining time series forecasts may aid in improving overall forecast accuracy in energy demand modeling. Using Turkey's daily household electricity consumption data from 2016 to 2021, this study thus investigates the potential benefits of combination models. Based on Schulze voting considering various performance metrics, the generated results indicate that combining forecasts from different traditional statistical models can result in better-performing models; however, it is not always the case that they outperform traditional models. In addition, static combination weights may not be suitable for wide forecast horizons. Potential avenues for future research include applying the investigated models to various time scales, such as hourly, weekly, monthly, or quarterly data for electricity consumption, to gain a deeper understanding of their performance and flexibility.

9.1 Introduction

The world's energy needs are constantly changing. Our economies depend on energy for a variety of purposes such as heating and cooling buildings, manufacturing processes, and transportation. Accurate and reliable energy demand forecasting

E. Isikli (✉) · S. Serdar Asan
Industrial Engineering Department, Faculty of Management, Istanbul Technical University,
Istanbul, Türkiye
e-mail: isiklie@itu.edu.tr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
G. Kayakutlu, M. Ö. Kayalica (eds.), *Decision Making Using AI in Energy and Sustainability*, Applied Innovation and Technology Management,
https://doi.org/10.1007/978-3-031-38387-8_9

153

is the basis of sustainable planning, operation, and control of the systems that generate and use these energy sources. Therefore, there is a growing interest in developing models to accurately predict energy demand. A plethora of papers suggest that current approaches adopted for this purpose can be basically classified into two categories: causal modeling and time-series modeling. The former category attempts to forecast energy demand by adopting an analytical approach with the use of various factors such as economic activity, weather, and population, which are, in turn, influenced by technology, policy, and geopolitics [1, 2]. Models with more information or factors may be better at explaining the behavior of energy demand; however, since they may also require the prediction of these factors accurately, energy demand forecasting through causal models can become highly complex and tedious. On the other hand, time series forecasting has been a key mechanism for modeling energy consumption due to its relative flexibility, and it appears to be expanding its sphere of influence in the age of big data [3–5]. Research on energy demand forecasting adopting traditional time series techniques is quite comprehensive [1–3, 6, 7]; however, there is also a growing use of computationally intelligent techniques such as artificial neural networks, fuzzy logic, and genetic algorithms [8]. Considerable literature has recently grown up around the theme of hybrid or multi-stage models that adopt the combined use of multiple techniques to improve model performance in terms of accuracy, speed, and complexity [9–14]. Combining time series forecasts, which can also be considered as a hybrid modeling approach, is not a new concept, and compared to sticking to a single forecasting model, it has been shown to be superior regarding accuracy and speed as complexity increases [6, 7, 9, 14, 15]. Time series that are highly volatile with nonstationary means and variances or that are relatively shorter, for instance, may benefit from the combination of forecasts from alternative models whether they are computationally intelligent or not. However, research to date has mostly concentrated on combining only a limited number of time series models of the same type through a single forecast combination method. This chapter thus investigates how to properly combine some of the most prominent time series models to improve overall forecast accuracy in energy demand modeling considering several model selection criteria. Data on daily electricity consumption in Turkey, covering a period of 6 years, was used to generate predictions through various time series forecasting techniques. These base models were trained on the first 80% of the observations, and their predictions were then combined through different methods to attain an overall forecasting performance that is higher than theirs. Performance metrics such as MAPE, hypothesis testing based on the DM statistic, and graphical inspection were used to evaluate the performance of alternative models. The contribution of the present study lies in its extensive use of time series models for forecast generation and its consideration of various ways to combine them. In addition, based on a variety of performance metrics, Schulze voting was employed to compare the alternative models with each other to decide which of them performed better. Understanding how the past influences the future, planners and policymakers can evidently design more efficient power grids, identify the most cost-effective methods of energy production, and optimize the use of energy resources. The investigation in the current study thus

contributes to such efforts flagging the importance of forecast combination in time series modeling.

9.2 Methodology

9.2.1 Model Selection

Deciding which time series model outperforms its contenders is not an easy task. In this subsection, we focus on the three most commonly used ways for model evaluation: performance metrics, hypothesis testing, and graphical inspection. The evaluation criteria to distinguish between alternative models should primarily be practical and efficient.

9.2.1.1 Performance Metrics

The most commonly used performance metrics that are given in Table 9.1 were written by adopting the in-sample notation whenever possible. However, we should note that the out-of-sample performance of a forecasting model, which depends on the number of observations used to generate forecasts and the forecast horizon, should essentially be used to distinguish between alternative models. Hence, these metrics should be modified accordingly when comparing the out-of-sample performances of the candidates. Most of the following metrics have adaptations for seasonal data, but here we provide only their generic versions. Note also that n denotes the in-sample size and h denotes the forecast horizon.

MAE, medAE, and RMSE are called scale-dependent metrics; MAPE, medAPE, maxAPE, sMAPE, and AMAPE are called percentage-error metrics; RelRMSE and TIC are called relative measures; MASE is called a scale-free error. The model with lower error measures (e.g., MAE, MAPE, maxAPE, RMSE) is preferred over the other alternatives. As the predicted values become closer to the actual values, all the error measures approach zero, indicating better forecasting accuracy. The same applies to FB and PBIAS; however, when FB is positive and PBIAS is negative, the model is said to underestimate the actual values [17]. MAPE is affected by the presence of relatively lower actual values [18]. TIC can take values within the interval $[0, 1]$. MASE, which compares the forecasting performance of a candidate model against the naive model, is particularly useful when the time series is intermittent.

9.2.1.2 Hypothesis Testing

To decide whether a model outperforms another, the DM test that was proposed by Diebold and Mariano [21] has been widely used [22]. The procedure compares the

Table 9.1 The most commonly used performance metrics

Name	Mathematical formula
Mean absolute error (MAE)	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
Median absolute error (medAE)	$\text{median}_{i=1,2,\dots,n} y_i - \hat{y}_i $
Root mean squared error (RMSE)	$\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$
Mean absolute percentage error (MAPE)	$\frac{100}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right $
Median absolute percentage error (medAPE) [16]	$100 \cdot \text{median}_{i=1,2,\dots,n} \left \frac{y_i - \hat{y}_i}{y_i} \right $
Maximum absolute percentage error (maxAPE) [17]	$100 \cdot \max_{i=1,2,\dots,n} \left \frac{y_i - \hat{y}_i}{y_i} \right $
Symmetric mean absolute percentage error (sMAPE) [10, 15]	$\frac{200}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{ y_i + \hat{y}_i }$
Average mean absolute percentage error (AMAPE) [14]	$\frac{100}{n} \sum_{i=1}^n \left[\frac{ y_i - \hat{y}_i }{\frac{1}{n} \sum_{i=1}^n y_i} \right]$
Relative root mean squared error (RelRMSE) [10]	$\frac{\text{RMSE}}{\frac{1}{n} \sum_{i=1}^n y_i}$
Theil's inequality coefficient (TIC) [18, 19]	$\frac{\sqrt{\frac{1}{h} \sum_{i=n+1}^{n+h} (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{h} \sum_{i=n+1}^{n+h} y_i^2 + \frac{1}{h} \sum_{i=n+1}^{n+h} \hat{y}_i^2}}$
Mean absolute scaled error (MASE) [20]	$\frac{\sum_{i=n}^{n+h} \left \frac{y_i - \hat{y}_i}{h} \right }{\sum_{i=2}^n \left \frac{y_i - y_{i-1}}{n-1} \right }$
Fractional bias (FB) [17]	$2 \cdot \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{\sum_{i=1}^n (y_i + \hat{y}_i)}$
Percent bias (PBIAS) [17]	$\frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{\sum_{i=1}^n y_i}$
Nash–Sutcliffe (NS) criterion [9]	$1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n \left(y_i - \frac{1}{n} \sum_{i=1}^n y_i \right)^2}$
Correlation coefficient (r)	$\frac{\text{Cov}(y_i, \hat{y}_i)}{\sqrt{\text{Var}(y_i) \text{Var}(\hat{y}_i)}}$

forecasting accuracy of two models using the following hypotheses based on the loss differential $\Delta_t = L(\varepsilon_t^A) - L(\varepsilon_t^B)$:

$$\begin{aligned} H_0 : & E[\Delta_t] = 0 \quad \text{for all } t \\ H_1 : & E[\Delta_t] \neq 0 \quad \text{for all } t \end{aligned}$$

where $L(\cdot)$ is the loss function that can be an $\mathcal{L}_1$ or an $\mathcal{L}_2$ norm and the DM statistics follows the standard normal distribution. The one-sided version of this test is also possible.

9.2.1.3 Graphical Inspections

Scatter plots of the actual values versus the predicted values or the residuals provide a simple way to comment on the model fit. Box plots and normal probability plots of the residuals generated by alternative models help differentiate between their performances. ACF and PACF plots of the residuals are particularly convenient to use in the ARIMA-type models to understand whether additional AR or MA terms are necessary or if the assumption of residual independence is violated.

9.2.2 Combining Time Series Forecasts

When working with a time series, it is a common practice to create alternative models to generate forecasts. The practitioner selects the best-performing one (s) among these models based on some predetermined criteria (see Table 9.1). However, these alternative models can also be combined to create some sort of synergy. Combining forecasts is not a recent methodology, with the first study dating back to the late 1960s [23]. Several methods currently exist in the related literature for combining forecasts produced by alternative time series models of the same type [24] or the different types [25]. Combined models were claimed to perform better than the worst-performing models and sometimes even outperform the best-performing model [26, 27] and they were especially recommended in the presence of outliers [28].

Several forecast combination methods have been proposed to achieve a lower overall forecast error:

1. *Mean combination*: The arithmetic, geometric, or harmonic mean of the predictions across all alternative models [25, 29]. The trimmed or winsorized means should be preferred when outliers exist in the data.
2. *Median combination*: The median of the predictions across all alternative models [25, 29].
3. *Weighted combination*: The weighted average of the predicted values is taken based on the procedure explained in Newbold and Granger [30]. The weights are

determined based on the covariances between the multi-step ahead forecast errors of the alternative models. The optimal combination weights, w_i , that minimize the variance of the forecast errors are calculated as follows:

$$\mathbf{w} = [w_1 \ w_2 \ \cdots \ w_k]^T = \frac{\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}$$

where $\sum_{i=1}^k w_i = 1$ and $w_i \in (0, 1)$ for $i = 1, 2, \dots, k$, $\mathbf{S}$ is the covariance matrix estimated using the in-sample forecast errors, and $\mathbf{1}$ is a vector of 1's. Previously, Bates and Granger [31] had used the diagonal elements of $\mathbf{S}$ to determine the combination weights. More recently, Aiolfi and Timmermann [32] suggested ranking the alternative models first based on their in-sample performance, and then assigning each model a combination weight that is proportional to the multiplicative inverse of its rank:

$$w_i = \frac{\frac{1}{\text{rank}_i}}{\sum_{j=1}^k \frac{1}{\text{rank}_j}}$$

4. *Regression combination*: A constrained regression model is fitted to the time series with the predicted values from alternative models as explanatory variables [24, 29, 33]. The combining weights, which were estimated using the estimation data, must sum up to one to produce combined forecasts in the validation data. The unconstrained version is also possible, but it has as many disadvantages as its advantages. Elliott and Timmermann [34] suggested using a loss function that is based on the absolute values of the forecast errors rather than the squared errors.
5. *The aggregated forecast through exponential reweighting (AFTER)* was first proposed by Yang [35] to combine different forecasts which were later applied by Zou and Yang [36] to the case where forecasts were generated using the same type of models. It has been modified [37, 38] and applied to various cases [39] since its first introduction.

There are also other combination methods that are based on the eigenvectors of $\mathbf{S}$ and the Generalized Rolling Origin Evaluation Combination (GROEC) method that has been recently proposed as a weighted combination method in which the weights are determined through a loss function [40].

9.3 Results and Discussion

We conducted experiments with the time series of daily electricity consumption in Turkey between January 1, 2016 and December 31, 2021 (a total of 2192 observations) using the R open-source statistics software. Figure 9.1 depicts the original time series along with its three seasonal components: Season1 – weekly seasonality (frequency = 7), Season2 – Hijri annual seasonality (frequency = 354.37), and Season3 – Gregorian annual seasonality (frequency = 365.25). The series was split into a fitting sample (training set) of 1827 observations (5 years) and a forecasting sample (test set) of 365 observations (1 year). We first determined six base models and estimated them via **forecastHybrid** and **ATAforecasting** packages and then combined their results using **ForecastComb** and **AIafter** packages. All of the model orders were verified in several in-sample forecasting trials. Descriptions of all the models considered in this study are given in Tables 9.2 and 9.3. Once we estimated the model parameters using the training sample, we produced 365-step-ahead forecasts by the candidate models to evaluate their post-sample performances. Then, based on the performance metrics given in Table 9.4, we ranked all the models based on each metric and created rank ballots to employ Schulze voting [41]. The models were eventually ordered from the best-performing to the worst as follows: dynOLS, STLTM, AFTER, ATA, MEDIAN, THETA, AMEAN, ARIMA, TBATS, WMEAN,

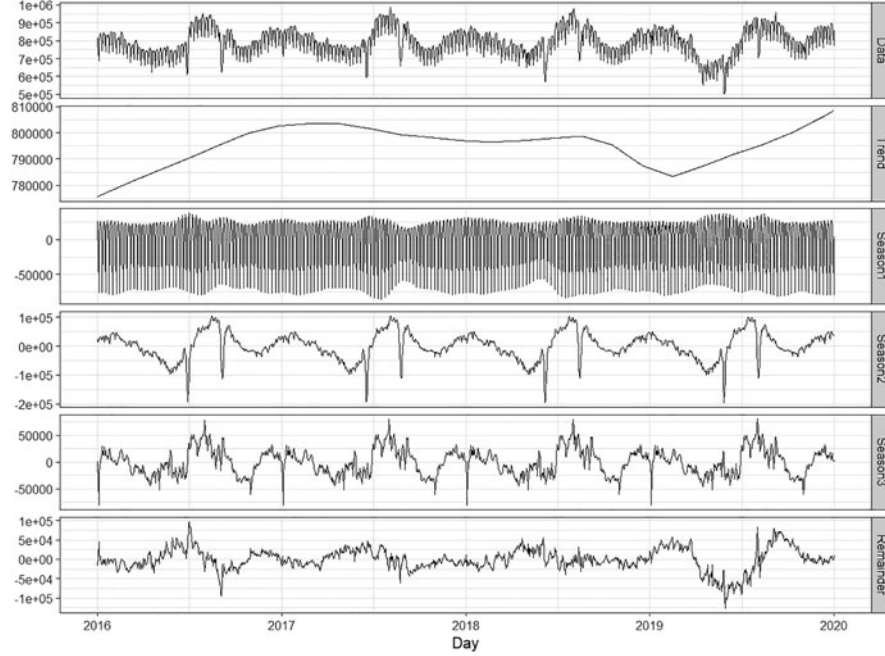


Fig. 9.1 Decomposition of Turkey electricity consumption time series

Table 9.2 Details on the base models

Abbreviation	Description and references
ARIMA	Non-seasonal Autoregressive Integrated Moving Average model with $p = 5$, $d = 1$, and $q = 2$.
NNETAR	Average of 20 networks, each of which is a Neural Network Autoregressive Model with 17 nodes in the hidden layer, 32 lagged inputs, and also the last observed values from the same season as inputs.
THETA	Theta model by Assimakopoulos and Nikolopoulos [42] with the smoothing parameter $\alpha = 0.7585$ and initial state $l = 725,128.24$.
STLM	STLM is a seasonal-trend decomposition procedure based on local regression smoothing (LOESS) that was developed by Cleveland et al. [43]. In the present study, a simple exponential smoothing with additive errors and multiple seasonality was constructed ($\alpha = 0.9686$, $l = 735,625.49$).
TBATS	TBATS is an exponential smoothing state space model developed by De Livera, Hyndman, and Snyder [44] for time series with multiple and complex seasonal patterns. In the abbreviation, B = Box-Cox transformation, A = ARMA errors, T = Trend, S = Seasonality, and the first T means that a trigonometric formulation was adopted to model seasonality. In the present study, the level parameter $\alpha = 0.0199175$ and the trend parameter $\beta = 0$. There are a total of six seasonal parameters, four AR coefficients, and one MA coefficient. The data was not Box-Cox transformed.
ATA	Proposed by Yapar et al. [45], the ATA method has been proclaimed to perform better than the other members of the family of exponential smoothing methods. The ATA model used in this study has additive errors, damped trend, and additive seasonality. Both the optimum level and trend parameters are equal to 55, and the optimum damped trend parameter equals 0.97.
AMEAN	Takes the arithmetic mean of the forecasts produced by the candidate models for each time period.
MEDIAN	Takes the median of the forecasts produced by the candidate models for each time period.
WEIGHTED	Takes the weighted average of the forecasts produced by the candidate models for each time period. The combination weights were determined based on the in-sample MASE.
BG	The weighted combination method proposed by Bates and Granger [31].
NG	The weighted combination method proposed by Newbold and Granger [30].
InvW	The weighted combination method proposed by Aiolfi and Timmermann [32].
OLS	The combination weights are simply the unconstrained beta parameters obtained by regressing forecasts from alternative models on the actual values.
LAD	The regression-based combination method proposed by Elliott and Timmermann [34].
dynOLS	The combination weights are found basically as in OLS, but they are allowed to vary by time.
AFTER	The Aggregated Forecast Through Exponential Reweighting method where the combination weights are non-linear.

Table 9.3 Combination weights

Abbreviation	ARIMA	THETA	NNETAR	STLM	TBATS	ATA
AMEAN	1/6	1/6	1/6	1/6	1/6	1/6
WEIGHTED	0.0855	0.0402	0.3257	0.2833	0.1925	0.0729
BG	0.0399	0.0109	0.4087	0.3841	0.1306	0.0258
InvW	0.1020	0.0680	0.4082	0.2041	0.1361	0.0816
LAD	−0.0150	−0.0249	0.7990	0.3897	−0.1738	0.0033
NG	−0.0108	−0.0061	0.7986	0.5427	−0.3288	0.0042
OLS	−0.0067	−0.0174	0.7951	0.5473	−0.3295	−0.0027
dynOLS	[−0.0323, 0.1359]	[−0.0576, 0.1259]	[−0.1240, 0.8544]	[0.4883, 0.9999]	[−0.7708, 0.1566]	[−0.0196, 0.1216]
AFTER	[0, 1/6]	[0, 1/6]	[0, 1]	[0, 1]	[0, 0.45]	[0, 1/6]

InvW, BG, NG, OLS, LAD, NNETAR. This ranking reveals several interesting results:

- dynOLS is the best-performing model both overall and among the combination models.
- AFTER is a promising alternative to dynOLS.
- STLM is the best-performing model among the base models and ATA is a promising alternative to it.
- MEDIAN, one of the simplest combination methods, ranked higher than four of the base models.
- dynOLS performed considerably better than both its static counterpart (OLS) and the other regression-based methods (BG, NG, and LAD).
- NNETAR is the worst-performing model both overall and among the base models.
- Regression-based methods, excluding dynOLS, performed worst among the combination models.

Next, we compared dynOLS with the second, third, and fourth best-performing models using the Diebold-Mariano test and graphical inspection. When dynOLS was tested against STLM, AFTER, and MEDIAN using a second-order loss-differential with $h = 1$, all of the one-sided null hypotheses were rejected. That is, dynOLS is significantly better performing than these models in terms of forecast accuracy. Finally, Fig. 9.2 depicts the actual testing dataset (blue line) and its forecast (red line) through the best base model and the three best combination models, and Fig. 9.3 illustrates how well the standardized residuals from each of these four models fit a proper normal distribution.

Table 9.4 Post-sample performances of the base and combination models

Model	MAE	medAE	RMSE	MASE ^a	MASE _a	MAPE	AMAPE	sMAPE	medAPE	maxAPE	PBIAS	FB	TIC	NSC	Corr
ARIMA	92,124	72,505	114,072	2.19	1.34	9.88%	10.29%	10.47%	8.82%	39.78%	0.07	0.09	0.07	-0.69	0.17
ATA	61,381	38,128	87,877	1.46	0.89	6.56%	6.86%	6.84%	4.48%	31.73%	0.04	0.05	0.05	0.00	0.51
NNETAR	207,737	216,233	238,622	4.93	3.02	22.72%	23.20%	26.58%	34.11%	75.03%	0.22	0.25	0.15	-6.39	0.13
STLM	54,266	46,338	68,405	1.29	0.79	5.98%	6.06%	6.20%	5.31%	43.39%	0.04	0.04	0.04	0.39	0.77
TBATS	94,555	83,170	108,439	2.24	1.37	10.54%	10.56%	11.29%	9.99%	34.28%	0.10	0.10	0.06	-0.53	0.75
THETA	74,107	47,696	103,332	1.76	1.08	8.86%	8.28%	8.17%	4.90%	30.62%	-0.07	-0.06	0.06	-0.39	0.42
AMEAN	76,745	68,346	91,332	1.82	1.12	8.29%	8.57%	8.69%	8.50%	27.25%	0.07	0.08	0.05	-0.08	0.71
AFTER	54,753	47,607	68,885	1.30	0.80	6.03%	6.12%	6.03%	5.47%	43.39%	0.04	0.04	0.04	0.38	0.76
BG	116,910	118,398	133,424	2.78	1.70	12.75%	13.06%	13.82%	15.75%	46.79%	0.12	0.13	0.08	-1.31	0.57
dynOLS	45,905	34,526	61,069	1.09	0.67	5.01%	5.13%	5.01%	3.77%	45.49%	0.03	0.03	0.03	0.52	0.78
InvW	116,288	112,970	133,194	2.76	1.69	12.64%	12.99%	13.68%	14.96%	37.59%	0.12	0.13	0.08	-1.30	0.51
LAD	167,884	174,620	195,072	3.99	2.44	18.31%	18.75%	20.78%	24.62%	75.79%	0.17	0.20	0.12	-3.94	0.24
MEDIAN	67,450	55,409	83,814	1.60	0.98	7.25%	7.53%	7.58%	6.52%	26.56%	0.06	0.07	0.05	0.09	0.77
NG	158,768	167,404	187,118	3.77	2.31	17.29%	17.73%	19.53%	23.41%	80.27%	0.16	0.18	0.11	-3.54	0.24
OLS	159,612	168,088	188,214	3.79	2.32	17.38%	17.83%	19.65%	23.60%	81.29%	0.16	0.19	0.11	-3.60	0.23
WMEAN	106,733	104,093	121,634	2.53	1.55	11.62%	11.92%	12.48%	13.49%	37.60%	0.11	0.12	0.07	-0.92	0.62

^aMASE was calculated by the formula given in Table 9.1, dividing the post-sample MAE of a model by the in-sample MAE of the naive model. Since the forecast horizon is wide, it was later modified as MASE^{*}, which replaces the value in the denominator with the out-sample MAE of the naive model that accepts the last observation in the training set as the forecast value for all the observations in the test set. Nevertheless, MASE values should be approached carefully and be used to compare the base and combination models with each other instead of comparing them with the naive model

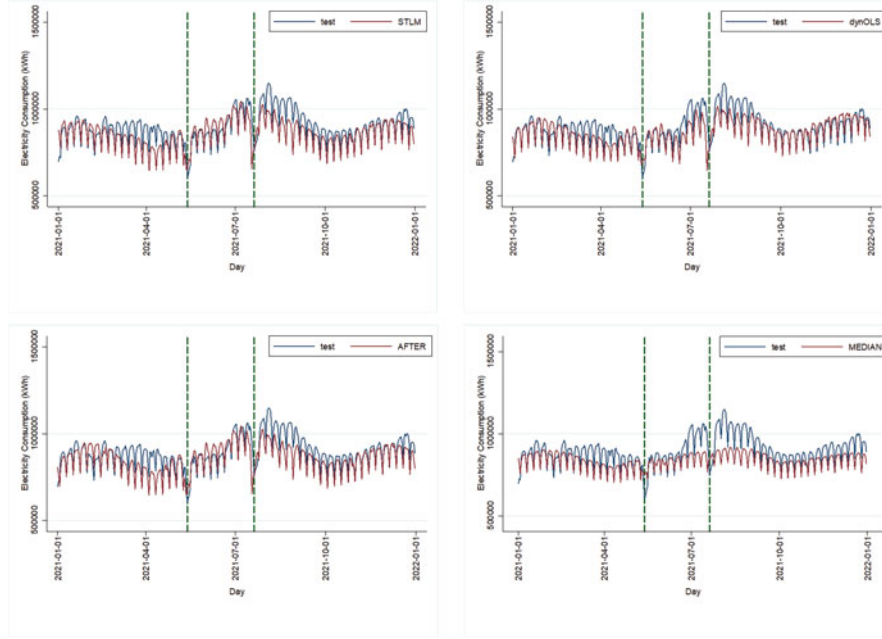


Fig. 9.2 Out-sample performances of the best base model and the three best combination models

9.4 Conclusion

Modern society is largely dependent on the supply of energy. To keep the lights on, the economy running, and the Internet humming, there is a clear need for better energy planning and management. Time series forecasting has the potential to improve the planning of sustainable energy systems by reducing the risk associated with long-term projections and the need for costly energy infrastructure upgrades. Recent developments in the field of computational intelligence have been leading researchers to modify traditional time series techniques or create hybrid models combining them with machine learning algorithms to improve prediction accuracy on complex data. Such approaches have the capability of improving overall forecast accuracy. Therefore, based on several performance metrics, the present study compares various ways to combine time series models using data on daily electricity consumption. In parallel with the related literature, our findings support combining the forecasts produced by different time series models. Even though it is not possible to claim that combination dominated the single (base) models considered, we realize that simple combination approaches (e.g., MEDIAN) can perform almost as well as their more sophisticated counterparts (e.g., AFTER). However, it is not recommended to work with static combination weights when the forecast horizon is wide.

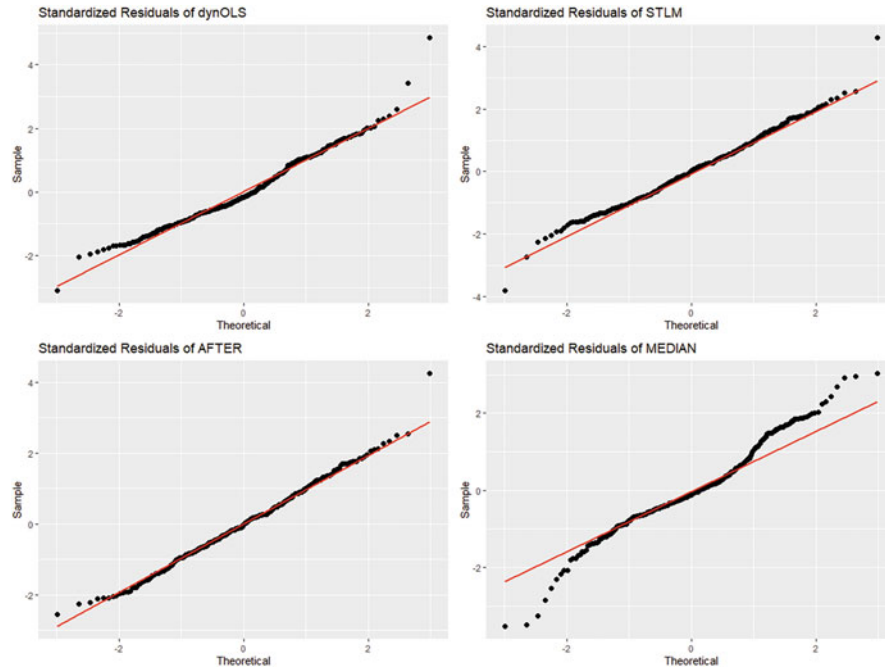


Fig. 9.3 Normal probability plots for the standardized residuals produced obtained from dynOLS, STLM, AFTER, and MEDIAN

As a possible avenue for future research, the investigated models can be applied to the cases of hourly, weekly, monthly, or quarterly data for electricity consumption to gain a better understanding of their performance and flexibility. Concerning sustainable energy production, transportation-related energy data or environmental data should also be investigated using the candidate models covered in this study accompanied by recently proposed methods such as Prophet.

References

1. Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting - A review. *Renewable and Sustainable Energy Reviews*, 16.
2. Ghalekhondabi, I., Ardjmand, E., Weckman, G. R., & Young, W. A. (2017). An overview of energy demand forecasting methods published in 2005–2015. *Energy Systems*, 8, 411–447. <https://doi.org/10.1007/S12667-016-0203-Y/TABLES/8>
3. Verwiebe, P. A., Seim, S., Burges, S., et al. (2021). Modeling energy demand: A systematic literature review. *Energies (Basel)*, 14, 7859. <https://doi.org/10.3390/EN14237859>
4. Makridakis, S., Spiliotis, E., Assimakopoulos, V., et al. (2021). The M5 uncertainty competition: Results, findings and conclusions. *International Journal of Forecasting*. <https://doi.org/10.1016/J.IJFORECAST.2021.10.009>
5. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*. <https://doi.org/10.1016/J.IJFORECAST.2021.11.013>
6. Mosavi, A., Salimi, M., Ardabili, S. F., et al. (2019). State of the art of machine learning models in energy systems, a systematic review. *Energies*, 12, 1301. <https://doi.org/10.3390/EN12071301>
7. Debnath, K. B., & Mourshed, M. (2018). Forecasting methods in energy planning models. *Renewable and Sustainable Energy Reviews*, 88, 297–325. <https://doi.org/10.1016/j.rser.2018.02.002>
8. Roozbeh Nia, A., Awasthi, A., & Bhuiyan, N. (2021). Industry 4.0 and demand forecasting of the energy supply chain: A literature review. *Computers and Industrial Engineering*, 154, 10.1016/j.cie.2021.107128.
9. Anele, A. O., Hamam, Y., Abu-Mahfouz, A. M., & Todini, E. (2017). Overview, comparative assessment and recommendations of forecasting models for short-term water demand prediction. *Water (Switzerland)*, 9. <https://doi.org/10.3390/w9110887>
10. da Silva, R. G., Ribeiro, M. H. D. M., Mariani, V. C., & Coelho, L. d. S. (2020). Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables. *Chaos, Solitons and Fractals*, 139. <https://doi.org/10.1016/j.chaos.2020.110027>
11. Guo, Y., Xu, Y. P., Sun, M., & Xie, J. (2021). Multi-step-ahead forecast of reservoir water availability with improved quantum-based GWO coupled with the AI-based LSSVM model. *Journal of Hydrology*, 597. <https://doi.org/10.1016/j.jhydrol.2020.125769>
12. Ho, W. T., & Yu, F. W. (2021). Predicting chiller system performance using ARIMA-regression models. *Journal of Building Engineering*, 33. <https://doi.org/10.1016/j.jobbe.2020.101871>
13. Ordóñez C, Sánchez Lasheras F, Roca-Pardiñas J, Juez FJ de C (2019) A hybrid ARIMA–SVM model for the study of the remaining useful life of aircraft engines. *Journal of Computational and Applied Mathematics* 346: <https://doi.org/10.1016/j.cam.2018.07.008>.
14. Wu, L., & Shahidepour, M. (2010). A hybrid model for day-ahead price forecasting. *IEEE Transactions on Power Systems*, 25. <https://doi.org/10.1109/TPWRS.2009.2039948>
15. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecast methods. *International Journal of Forecasting*, 36. <https://doi.org/10.1016/j.ijforecast.2019.04.014>
16. Tyralis, H., & Papacharalampous, G. (2017). Variable selection in time series forecasting using random forests. *Algorithms*, 10. <https://doi.org/10.3390/a10040114>
17. Ng, K. Y., & Awang, N. (2018). Multiple linear regression and regression with time series error models in forecasting PM10 concentrations in Peninsular Malaysia. *Environmental Monitoring and Assessment*, 190. <https://doi.org/10.1007/s10661-017-6419-z>
18. Fang, T., & Lahdelma, R. (2016). Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Applied Energy*, 179. <https://doi.org/10.1016/j.apenergy.2016.06.133>

19. Faghih, S. A. M., Gholipour, Y., & Kashani, H. (2021). Time series analysis framework for forecasting the construction labor costs. *KSCE Journal of Civil Engineering*, 25. <https://doi.org/10.1007/s12205-021-1489-4>
20. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS One*, 13. <https://doi.org/10.1371/journal.pone.0194889>
21. Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13. <https://doi.org/10.1080/07350015.1995.10524599>
22. Halužan, M., Verbič, M., & Zorič, J. (2020). Performance of alternative electricity price forecasting methods: Findings from the Greek and Hungarian power exchanges. *Applied Energy*, 277. <https://doi.org/10.1016/j.apenergy.2020.115599>
23. Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5. [https://doi.org/10.1016/0169-2070\(89\)90012-5](https://doi.org/10.1016/0169-2070(89)90012-5)
24. Tselentis, D. I., Vlahogianni, E. I., & Karlaftis, M. G. (2015). Improving short-term traffic forecasts: To combine models or not to combine? *IET Intelligent Transport Systems*, 9. <https://doi.org/10.1049/iet-its.2013.0191>
25. Milunovich, G. (2020). Forecasting Australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting*, 39. <https://doi.org/10.1002/for.2678>
26. Oh, C. O., & Morzuch, B. J. (2005). Evaluating time-series models to forecast the demand for tourism in Singapore: Comparing within-sample and postsample results. *Journal of Travel Research*, 43. <https://doi.org/10.1177/0047287505274653>
27. Song, H., & Li, G. (2008). Tourism demand modelling and forecasting-a review of recent research. *Tourism Management*, 29. <https://doi.org/10.1016/j.tourman.2007.07.016>
28. Beyaztas, U., Shang, H. L., & Yaseen, Z. M. (2021). A functional autoregressive model based on exogenous hydrometeorological variables for river flow prediction. *Journal of Hydrology*, 598. <https://doi.org/10.1016/j.jhydrol.2021.126380>
29. Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting*, 27. <https://doi.org/10.1016/j.ijforecast.2010.05.019>
30. Newbold, P., & Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society, Series A*, 137. <https://doi.org/10.2307/2344546>
31. Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20, 451–468. <https://doi.org/10.1057/JORS.1969.103>
32. Aiolfi, M., & Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135. <https://doi.org/10.1016/j.jeconom.2005.07.015>
33. Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3. <https://doi.org/10.1002/for.3980030207>
34. Elliott, G., & Timmermann, A. (2004). Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics*, 122. <https://doi.org/10.1016/j.jeconom.2003.10.019>
35. Yang, Y. (2004). Combining forecasting procedures: Some theoretical results. *Econometric Theory*, 20. <https://doi.org/10.1017/S0266466604201086>
36. Zou, H., & Yang, Y. (2004). Combining time series models for forecasting. *International Journal of Forecasting*, 20. [https://doi.org/10.1016/S0169-2070\(03\)00004-9](https://doi.org/10.1016/S0169-2070(03)00004-9)
37. Wei, X., & Yang, Y. (2012). Robust forecast combinations. *Journal of Econometrics*, 166. <https://doi.org/10.1016/j.jeconom.2011.09.035>
38. Qian, W., Rolling, C. A., Cheng, G., & Yang, Y. (2022). Combining forecasts for universally optimal performance. *International Journal of Forecasting*, 38. <https://doi.org/10.1016/j.ijforecast.2021.05.004>

39. Borthakur, P., & Goswami, B. (2020). Short term load forecasting: A hybrid approach using data mining methods. In *2020 international conference on emerging frontiers in electrical and electronic technologies, ICEFEET 2020*.
40. Fiorucci, J. A., & Louzada, F. (2020). GROEC: Combination method via generalized rolling origin evaluation. *International Journal of Forecasting*, 36. <https://doi.org/10.1016/j.ijforecast.2019.04.013>
41. Schulze, M. (2011). A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method. *Social Choice and Welfare*, 36. <https://doi.org/10.1007/s00355-010-0475-4>
42. Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting. *International Journal of Forecasting*, 16. [https://doi.org/10.1016/S0169-2070\(00\)00066-2](https://doi.org/10.1016/S0169-2070(00)00066-2)
43. Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on LOESS (with discussion). *Journal of Official Statistics*, 6.
44. de Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106. <https://doi.org/10.1198/jasa.2011.tm09771>
45. Yapar, G., Selamlar, H. T., Capar, S., & Yavuz, I. (2019). ATA method. *Haceteppe Journal of Mathematics and Statistics*, 48. <https://doi.org/10.15672/hujms.461032>